

Clustering und unüberwachtes Lernen

Strukturen und Gruppen in Daten ohne vorgegebene Zielvariable entdecken

01 INTRO

Beim unüberwachten Lernen existiert keine vorgegebene Zielvariable. Stattdessen sollen Strukturen, Gruppen oder ungewöhnliche Beobachtungen direkt aus den Daten erkannt werden. Dadurch entstehen besondere Herausforderungen bei Modellwahl und Bewertung.

Der Kurs behandelt zentrale Verfahren des Clusterings und die systematische Interpretation ihrer Ergebnisse. Neben k-Means werden hierarchische und dichte-basierte Ansätze betrachtet und unterschiedliche Vorstellungen von Ähnlichkeit und Clusterstruktur verglichen.

DAUER

2 Tage

FORMAT

Präsenz · Live Online · Inhouse

VORKENNTNISSE

Python-Grundkenntnisse und grundlegendes Verständnis von Machine Learning

02 TRAININGSPROGRAMM

Inhalte

Unüberwachtes Lernen

- Lernen ohne Labels
- Exploration versus Modellierung
- typische Anwendungen
- Cluster
- Anomalien
- Grenzen der Interpretation

Ähnlichkeit und Distanz

- euklidische Distanz
- weitere Distanzmaße
- Skalierung
- Feature-Auswahl
- Bedeutung unterschiedlicher Dimensionen

Daten für Clustering vorbereiten

- fehlende Werte
- Skalierung
- kategoriale Merkmale
- Ausreißer
- redundante Features
- hohe Dimensionalität

k-Means

- Clusterzentren
- Zuordnung
- iterative Optimierung
- Initialisierung
- Anzahl der Cluster
- Grenzen des Verfahrens

Anzahl der Cluster bestimmen

- Inertia
- Elbow Method
- Silhouette Score
- Stabilität
- fachliche Interpretierbarkeit
- keine rein mechanische Entscheidung

Hierarchisches Clustering

- agglomerative Verfahren
- Linkage
- Dendrogramme
- Clusterhierarchien
- Auswahl von Schnittebenen

Dichtebasierte Verfahren

- DBSCAN
- Kerne und Nachbarschaften
- Noise
- Parameter
- nicht-kugelförmige Cluster
- Stärken und Grenzen

Cluster visualisieren

- Projektionen
- zweidimensionale Darstellungen
- Clusterprofile
- Verteilungen
- Grenzen visueller Interpretation

Cluster interpretieren

- charakteristische Merkmale
- Profile
- Gruppenvergleiche
- fachliche Bedeutung
- Post-hoc-Erklärungen
- Vorsicht vor Überinterpretation

Qualität unüberwachter Modelle

- interne Metriken
- externe Information
- Stabilität
- Sensitivität
- fachliche Validierung
- Grenzen objektiver Bewertung

Weitere Verfahren einordnen

- Anomalieerkennung
- Gaussian Mixtures
- Dimensionsreduktion
- Topic- und Latent-Strukturen im Überblick
- Auswahl nach Fragestellung

Praktisches Clustering-Projekt

- Daten vorbereiten
- Verfahren auswählen
- Parameter untersuchen
- Modelle vergleichen
- Cluster profilieren
- Ergebnisse fachlich interpretieren

03 DURCHFÜHRUNG

Rahmen und Organisation

Dauer	2 Tage
Durchführung	Präsenz oder Live Online, vorzugsweise als Inhouse-Schulung
Schwerpunkt	Inhouse-Schulungen für Unternehmen und Teams
Sprache	Deutsch, englische Fachbegriffe und Dokumentation
Unterlagen	Kursunterlagen, Beispiele und Übungsaufgaben
Technik	Eigener Rechner mit Python, Jupyter und scikit-learn

Inhouse-Schulungen

Eigene Datenbestände und Segmentierungsfragen können in die Übungen einbezogen werden.

04 ZIELGRUPPE

Für wen ist der Kurs gedacht?

Data Analysts, Entwickler und Data Scientists, die Strukturen in nicht gelabelten Daten untersuchen möchten.

Voraussetzungen

Grundkenntnisse in Python, Pandas und Machine Learning.

05 LERNZIELE

Was Sie nach dem Kurs können

Die Teilnehmer können Clustering-Probleme formulieren, verschiedene Verfahren einsetzen und vergleichen sowie Clusterergebnisse kritisch bewerten und interpretieren.

06 FAQ

Häufige Fragen

Woher weiß ich, wie viele Cluster existieren?

Das lässt sich häufig nicht eindeutig aus einer einzelnen Kennzahl bestimmen. Der Kurs behandelt statistische Kriterien, Stabilität und fachliche Interpretation gemeinsam.

Wird Dimensionsreduktion behandelt?

Sie wird dort eingesetzt, wo sie für Clustering und Visualisierung relevant ist. Der Kurs zu Feature Engineering behandelt sie systematisch.

Ist Clustering immer objektiv?

Nein. Ergebnis und Interpretation hängen unter anderem von Features, Skalierung, Distanzmaß und Verfahren ab.

UC IT Service · Ulrich Cuber · kontakt@uc-it.de

<https://www.uc-it.de/coaching/clustering/>